

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Welcome to this four-day, instructor-led training on *The Machine Learning Pipeline on AWS*.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

AWS Training and Certification

Module 0: Introduction



DO NOT COPY
amipandit@deloitte.com

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Logistics



aws training and certification

- Parking
- Facility:
 - Emergency exits
 - Fire alarm protocol
 - Security
- Breaks and lunch
- Food
- Cellular phones
- Student materials: Gilmore

© 2022 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Lab setup & books: One time steps



Resources you will need during this course

| | |
|---|---|
|  QWIKLABS |  eVantage |
| Lab work Register for Qwiklabs https://aws.qwiklab.com | Student manuals Register for Gilmore https://www.gilmore.ca/eVantageRegistration |
| Download bookshelf OR read online Download: https://support.vitalsource.com/hc/en-us Online: https://evantage.gilmoreglobal.com/ | |

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

Register for Gilmore

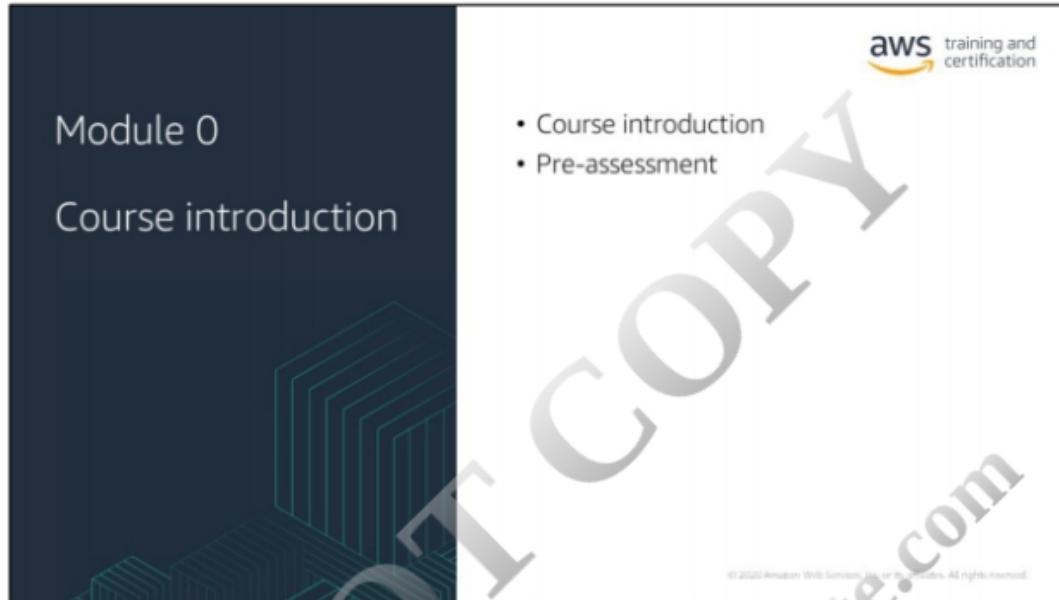
<https://www.gilmore.ca/eVantageRegistration>

Download bookshelf OR read online

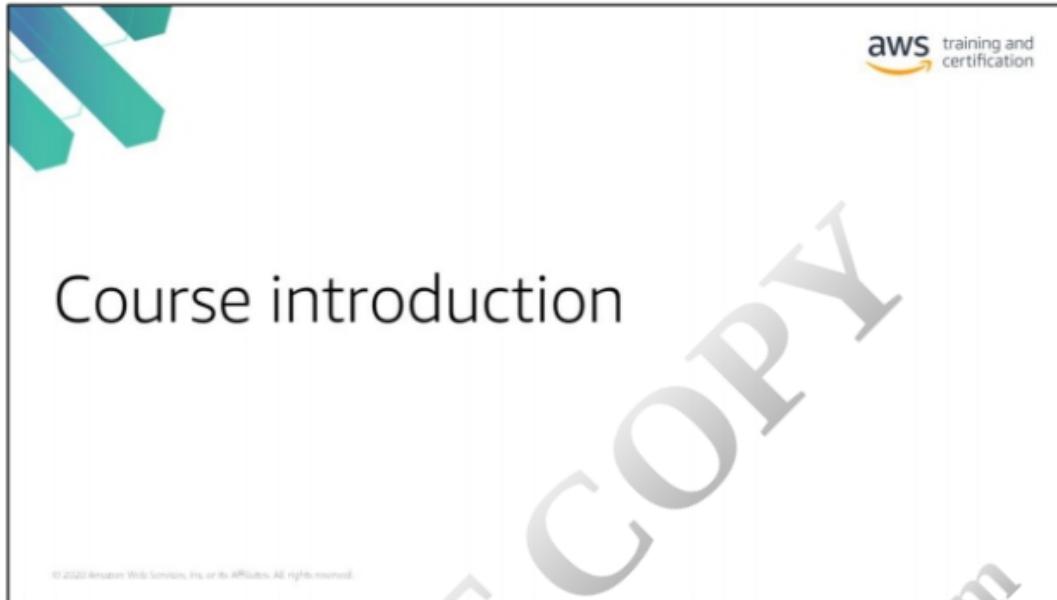
Download: <https://support.vitalsource.com/hc/en-us>

Online: <https://evantage.gilmoreglobal.com/>

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



© 2022 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Course description

aws training and certification



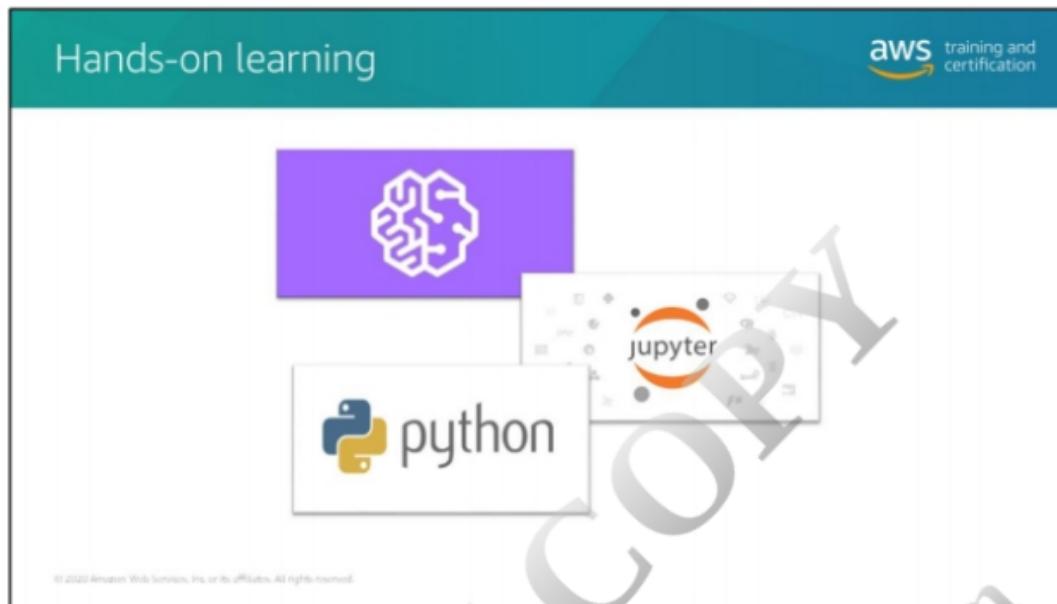
Learn and apply:

- Machine learning (ML) pipeline
- Amazon SageMaker

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

This course is designed to give you practical experience with machine learning. You'll learn how to use the ML pipeline to guide you through practical ML problems, and use Amazon SageMaker to build ML solutions.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



This course takes a hands-on, project-based approach to teaching you machine learning. You will be working in Amazon SageMaker and Jupyter notebooks, and you will use Python and some common Python ML libraries. However, while experience with these tools will accelerate your learning over the next four days, they are not prerequisites for this course.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Course prerequisites



Basic knowledge of Python:

- How to create functions, list comprehensions, dictionaries, AWS Lambda functions
- How to import libraries using pip

Basic understanding of the AWS Cloud:

- AWS Cloud infrastructure
- How to navigate Amazon S3 and Amazon CloudWatch

Baseline understanding of how to write code cells and Markdown cells in a **Jupyter** notebook environment

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

To get the most out of this course, you should have:

- Basic knowledge of Python
- A basic understanding of cloud computing on AWS
- A baseline understanding of working in a notebook environment

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

What is your background?

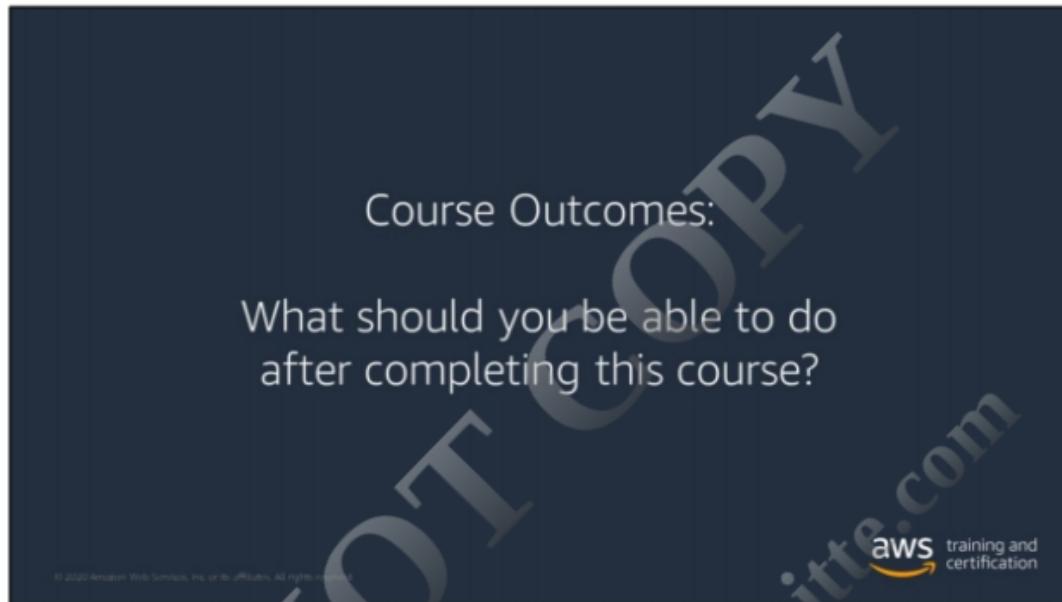
Introduce yourself and talk about your interactions with AWS and ML.

- *What do you do?*
- *How long have you been using AWS?*
- *What is your experience with ML?*
- *What do you hope to get out of this course?*

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

While AWS Training and Certification courses are designed with a particular audience in mind, students come with a variety of backgrounds, professional experiences, and learning goals. Take a minute to share with the class a little about yourself and what you want to get out of the course.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



As mentioned, this course is designed to walk you through the ML pipeline and AWS tools, such as Amazon SageMaker, that you will need to build, train and deploy ML solutions in the cloud. More specifically, by the end of the course, you should be able to:

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



- 1) Select and justify the appropriate ML approach for a given business problem.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



- 2) Build, train, evaluate, deploy, and fine-tune a machine learning model in Amazon SageMaker.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



- 3) Apply the steps of the ML pipeline to solve a specific business problem.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



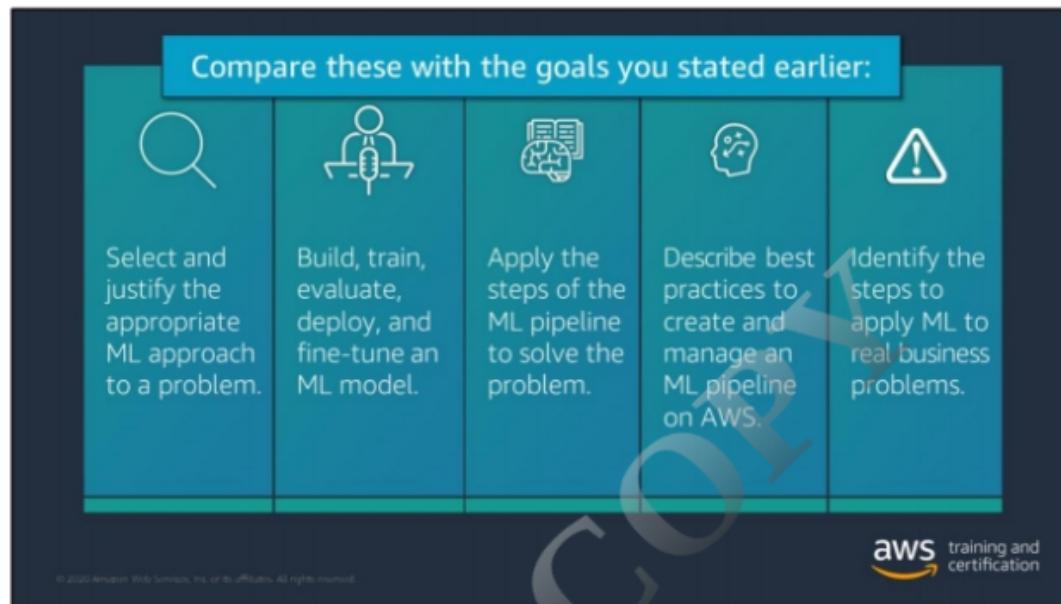
- 4) Describe some of the best practices for designing scalable, cost-optimized, reliable, and secure ML pipelines in AWS.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



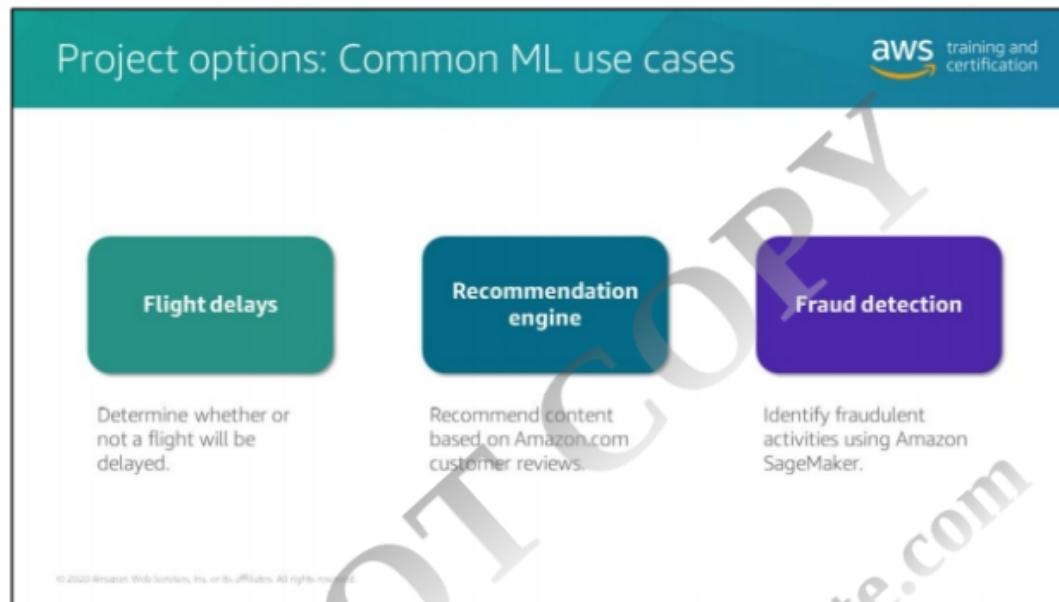
- 5) Identify the steps needed to apply machine learning to meet real-life business problems of interest after the course is complete.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Are these learning outcomes aligned with your goals for this course? What could you do to ensure that you meet your goals while also meeting these outcomes?

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



The slide is titled "Project options: Common ML use cases" and features the AWS training and certification logo. It lists three projects: Flight delays, Recommendation engine, and Fraud detection, each with a brief description.

| Project | Description |
|-----------------------|---|
| Flight delays | Determine whether or not a flight will be delayed. |
| Recommendation engine | Recommend content based on Amazon.com customer reviews. |
| Fraud detection | Identify fraudulent activities using Amazon SageMaker. |

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

To enable you to fulfill these learning outcomes and to make sure you get hands-on experience with machine learning, the course is centered around three different projects, one of which you will choose to work on throughout the four days. Each project represents a business problem that offers opportunities to us ML as a solution. Over four days, you will take an iterative approach to working on your project, slowly building, training, and deploying an ML model as a solution to your business scenario. You will do this by working directly in Jupyter notebooks via an Amazon SageMaker Notebook Instance. You will use the Python programming language in the notebooks.

This course takes this project-based approach, rather than a lecture-based approach, to give you practical experience using ML concepts and tools. Later today, you will get to familiarize yourself with each project. Once you fully understand them, you will choose one to work on for the remainder of the course. Let's briefly review the projects now.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Project details

Flight delays

Determine whether or not a flight will be delayed.

- You work for a travel booking website that seeks to improve the customer experience for people whose flights were delayed.
- The company wants to let customers know, when they book their flights, if their flights will be delayed due to weather.
- You need to use ML to identify whether a flight will be delayed due to weather.

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Project details

aws training and certification

Recommendation engine

Recommend content based on Amazon.com customer reviews.

- You work for a startup that delivers on-demand video streaming services to users.
- The company wants to introduce recommendations based on each user's viewing history.
- You need to use ML to create a recommendation engine.

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Project details

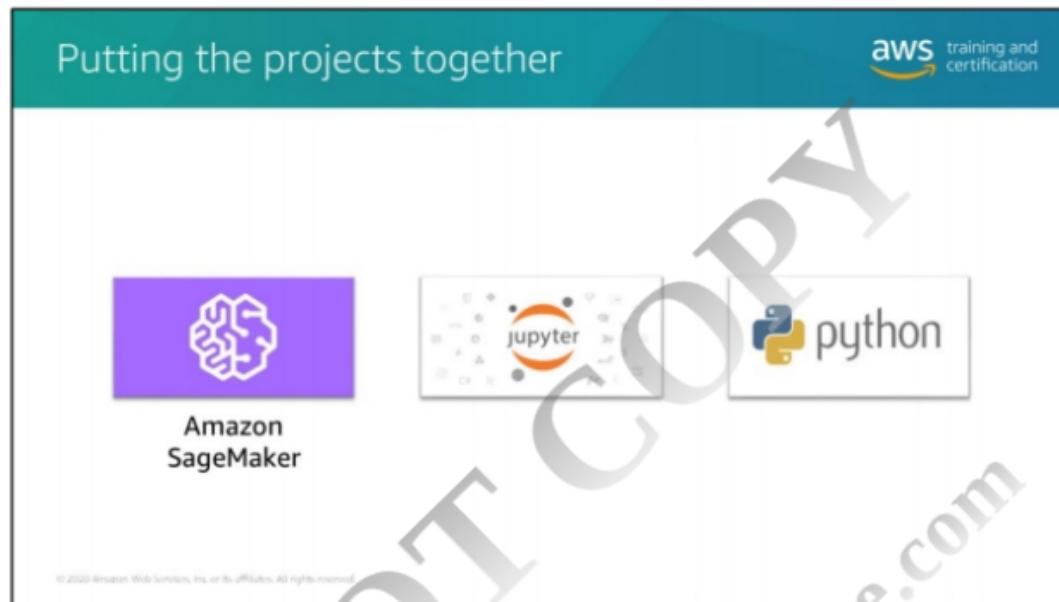
Fraud detection

Identify fraudulent activities using Amazon SageMaker.

- You work for a multinational bank.
- Over the last few months, there has been a significant uptick in the number of customers experiencing credit card fraud.
- You need to use ML to identify fraudulent credit card transactions before they have a larger impact on your company.

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



To conduct the project work for this course, you'll use a handful of tools that enable ML workloads—specifically, Amazon SageMaker, Python, and Jupyter notebooks.

Amazon SageMaker is a cloud-based platform where you can build your ML models, train them, and deploy them. But why do we need a platform? Each phase of this process – build, train, and deploy – requires a different kind of compute infrastructure, a different approach to solving problems, a distinct set of skills and mathematical knowledge to accomplish, and much more. SageMaker provides all of the components used to address those things in one toolkit – so you don't have to stitch together various workflows.

Within Amazon SageMaker, you'll use Jupyter notebooks: environments that allow you to create and share documents, code, and reports, and they enable you to perform data visualizations in the same space. With the complex workflows involved in machine learning, this increases both productivity and collaboration.

You will also use Python within the Jupyter notebooks. While you don't have to use Python to conduct machine learning, you will use Python in this course because it is simpler for humans to read (which is necessary when dealing with complex algorithms) and because you can use more than one language in the same notebook (which is best when collaborating with other developers, data scientists, etc.)

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Additional resources

aws training and certification

Student guide

- All slides and notes
- Follow along during training
- Use as resource after training

Student project template

- Different section for each course module
- Write down important topics
- Keep track of questions and things you need clarification on
- Summarize in your own words
- Linked in the notes

Example project

- Different problem: Targeted direct marketing campaign
- You can refer to it to get a sense of how to answer the questions
- Linked in the notes

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

You should have access to this course's Student Guide. The Student Guide includes all of the slides and slide notes that will be presented to you over these four days. You can use the Student Guide to follow along with this training and revisit it after the training is complete.

Later on, you will also get access to a note-taking project template, which you can use to note important concepts and topics as they are being discussed. The template, which includes a different section for each course module, is structured in a way that also encourages you to write down questions and topics you may need some follow-up or clarification on. And the last part of each template section includes a place for you to write down a summary, in your own words, of the most important concepts learned in each module.

At AWS Training and Certification, we are not only subject-matter experts in technical domains such as machine learning, but we are also experts in learning design, understanding how people learn most effectively. One way to maximize learning is to ensure students are fully engaged and active participants in the training. The more you synthesize the information being presented to you and summarize the key concepts, the better you will retain the information. So we strongly encourage you to use this template to take notes during this course.

And lastly, you will be given access to an example project.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Student project template: <https://aws-tc-largeobjects.s3-us-west-2.amazonaws.com/ILT-TF-200-MLDWTS/Student+Project+Template.docx>

Example project: <https://aws-tc-largeobjects.s3-us-west-2.amazonaws.com/ILT-TF-200-MLDWTS/Example+Project.docx>

DO NOT COPY
amipandit@deloitte.com

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

| Course outline | | | |
|--|---|--|--|
| Day 1 | Day 2 | Day 3 | Day 4 |
| <ul style="list-style-type: none">• Mod 0: Introduction• Exercise: Pre-assessment• Mod 1: Intro to ML and the ML Pipeline• Mod 2: Introduction to Amazon SageMaker• Lab 1: Introduction to Amazon SageMaker• Exercise: Choose your project• Mod 3: Problem formulation• Exercise: Formulate your project's business problem | <ul style="list-style-type: none">• Checkpoint #1• Mod 4: Data preprocessing• Lab 2: Data preprocessing• Mod 5: Model training | <ul style="list-style-type: none">• Checkpoint #2• Mod 6: Model evaluation• Lab 3: Model Training and Evaluation• Exercise: Project presentations• Mod 7: Feature Engineering and Model Tuning | <ul style="list-style-type: none">• Checkpoint #3• Lab 4: Feature Engineering• Mod 8: Model deployment• Exercise: Final project share out• Exercise: Post-assessment• Mod 9: Course Wrap-up |

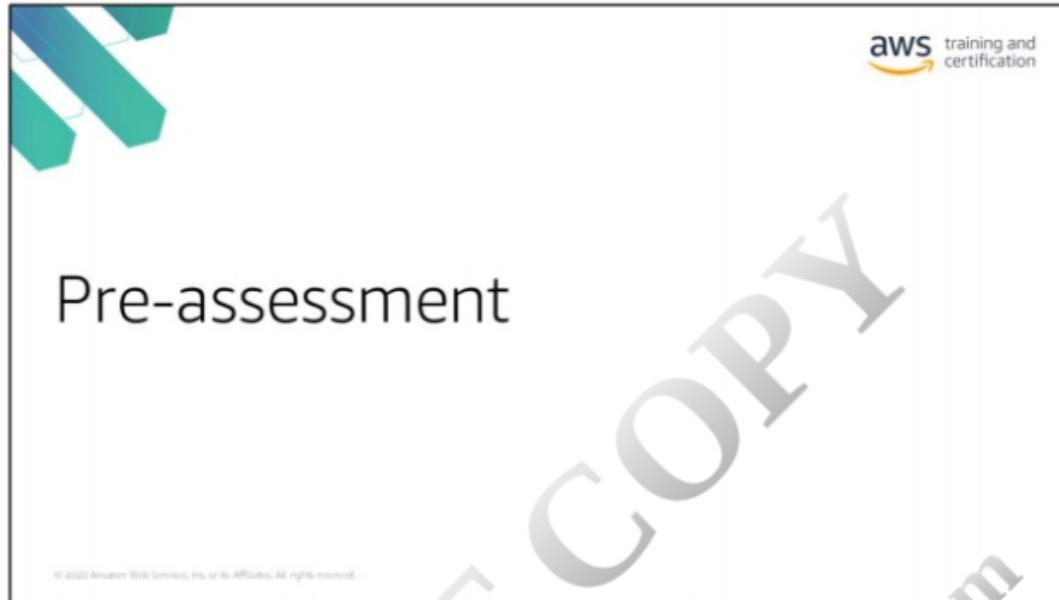
© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

Layered throughout your project work are instructor-led presentations, additional independent exercises, assessments and hands-on labs. The course starts on the first day with an introduction to machine learning and the ML pipeline, followed by an opportunity for you to dig further into the course projects and a chance to pick the project and team, if applicable, that is of most interest to you. Then we introduce you to Amazon SageMaker and begin discussing the pipeline. Then you'll learn about the different projects you may work on, along with problem formulation. On Day 2, you will focus on data preprocessing and model training. On Day 3, you will learn about model evaluation and then present your project, if you or your group would like to. The day will end with a discussion about feature engineering and model tuning. And finally, on Day 4, you will focus on productionizing your model and present your project to the class.

At the beginning of this training, you can take a test to assess how well you understand the core concepts in this course; another test is available to you at the end of the course, so that you can see what you've learned (which helps reinforce your learning). Similarly, you will take a short knowledge check at the very start of Days 2, 3 and 4, which helps you evaluate and reinforce key concepts from the previous days.

All of this will happen between 9 am and 5 pm over the next four days. Each day, there will be time set aside for breaks and lunch. Your instructor will specify when exactly those breaks will fall within the days, but in general, you will have one 15-minute break in the morning and one in the afternoon, as well as 60 minutes for lunch in the middle of the day.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



The slide features a dark blue background with a geometric pattern of light blue lines forming a stylized 'X' shape at the bottom. The title 'Pre-assessment overview' is centered in white text. In the top right corner is the AWS training and certification logo. Below the logo is a bulleted list of test details and covered domains. A small copyright notice is at the bottom right.

Pre-assessment overview

aws training and certification

- 25 questions
- 45-minute time limit
- Domains covered:
 - Introduction to Machine Learning and the ML Pipeline
 - Introduction to Amazon SageMaker
 - Problem Formulation and Data Understanding
 - Data Preprocessing
 - Model Training
 - Model Evaluation
 - Feature Engineering and Model Tuning
 - Model Deployment

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

Before jumping in, take 45 minutes to complete this 25-question pre-assessment test, the link to which can be found on the following slides. These 25 multiple choice questions cover a range of topics and concepts that you will become more familiar with over the next four days. Each question you encounter is categorized into one of the following course domains, most of which align to the main phases of the ML pipeline that you will learn about next. They are:

- Introduction to Machine Learning and the ML Pipeline
- Introduction to Amazon SageMaker
- Problem Formulation and Data Understanding
- Data Preprocessing
- Model Training
- Model Evaluation
- Feature Engineering and Model Tuning
- Model Deployment

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

Why take the pre-assessment?



- Start becoming familiar with topics covered in this course
- Understand your strengths and weaknesses in ML
- Create a benchmark to assess how much you learn

© 2022 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

DO NOT COPY
amipandit@deloitte.com

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.



Here is the link for the pre-assessment:

https://amazonmr.au1.qualtrics.com/ife/form/SV_0052a3eXy2Jnkbj

And note that at the end, you will see your overall score and the breakdown of your scores in each domain or topic area. Because you will take a similar post-assessment at the end of this course, you will not see your results for each question just yet, but you can use the overall and domain scores to give you a sense of how well you did and where your strengths and weaknesses may lie.

Printed by: amipandit@deloitte.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

